

**ANALYZING ACADEMIC ACHIEVEMENT OF CAS's STUDENTS
USING DATA MINING**

A thesis submitted to the College of Arts and Sciences in
partial fulfillment of the requirements for the degree of
Master of Science (Information Technology)
Universiti Utara Malaysia

by

Nor Asiah binti Abdul Rahman



KOLEJ SASTERA DAN SAINS
(College of Arts and Sciences)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certify that)

NOR ASIAH ABDUL RAHMAN
(89450)

calon untuk Ijazah
(candidate for the degree of) **MSc. (Information Technology)**

telah mengemukakan kertas projek yang bertajuk
(has presented his/her project paper of the following title)


ANALYZING ACADEMIC ACHIEVEMENT OF
CAS'S STUDENT USING DATA MINING

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project paper)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan
dan meliputi bidang ilmu dengan memuaskan.
(that the project paper acceptable in form and content, and that a satisfactory
knowledge of the field is covered by the project paper).

Nama Penyelia Utama
(Name of Main Supervisor): **ASSOC. PROF. FADZILAH SIRAJ**

Tandatangan
(Signature)

: 

Tarikh
(Date)

: 10-05-2009

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence by the Dean of Academic Affairs. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to

Division of Applied Sciences

College of Arts and Sciences

Universiti Utara Malaysia

06010 UUM Sintok

Kedah Darul Aman.

ABSTRACT (BAHASA MALAYSIA)

Sumber maklumat yang diperolehi daripada data pelajar dapat digunakan dan diproses menjadi pengetahuan yang berguna. Institusi pendidikan pada masa kini telah banyak mengumpulkan maklumat pelajar seperti maklumat akademik, demografik dan berkaitan dengan personality. Data yang telah dikumpulkan digunakan untuk meramal pencapaian akademik para pelajar. Kajian ini menggunakan responden yang terdiri daripada kalangan pelajar yang telah menamatkan pengajian pada tahun 2006, 2007 dan 2008. Dalam kajian ini, dua teknik perlombongan data telah digunakan untuk menganalisis dan membangunkan model klasifikasi untuk pencapaian akademik pelajar di Kolej Sastera dan Sains (CAS), Universiti Utara Malaysia (UUM). Pada peringkat awal, hubungan dan kolerasi di antara pencapaian purata nilai gred kumulatif (PNGK) dengan latarbelakang akademik, demografik, kelayakan masuk, penajaan, dan kemahiran sendiri dianalisis. Analisis menggunakan *Multinomial Logistic Regression* dan *Neural Network* menunjukkan jantina, kelayakan masuk, kelayakan bahasa (Bahasa Malaysia dan Inggeris), pendapatan keluarga, penajaan, kemahiran analitikal dan analisis serta kerja berkumpulan telah menyumbang utama kepada pencapaian akademik pelajar. Keputusan daripada kajian ini diharapkan agar dapat diguna pakai oleh pihak pengurusan CAS khususnya untuk membuat keputusan dan meramal pencapaian pelajar terkini dan akan datang.

ABSTRACT (ENGLISH)

Massive information can be collected from students' data in order to produce knowledge. The educational institutions collect students' data such as academic information, demographic, and personal traits. The data collected based on these variables used to predict the students' academic achievement. On this study, the respondents are students who have graduated within the period of six months in the year 2006, 2007 and 2008. Two data mining techniques for analyzing and building the classification model for students' achievement in College of Arts and Sciences (CAS), Universiti Utara Malaysia (UUM) are presented. Initially, the relationship and correlation between students' cumulative grade point average (CGPA) with academic background, demographic, entry qualification, sponsorship and interpersonal skills, students' achievement are analyzed. For model building purposes, final CGPA has been used as a target. The analysis conducted using Multinomial Logistic Regression and Neural Network found that, gender, entry qualification, language qualification (Bahasa Malaysia and English), family income, sponsorship, analytical and analysis skill as well as teamwork are all the best predictors contributed to students' performance. The result obtained through this study can be used to help the management of CAS to make certain decisions and to predict the outcome of current and future students.

ACKNOWLEDGEMENTS

Thanks to ALLAH S.W.T with His bless and mercy for giving me strength to accomplish my thesis and simultaneously completed my masters degree. With this opportunity, I would like to express my grateful to Universiti Utara Malaysia for the valuable chance by funding my education in master degree level.

I would like to express my gratitude to particular individuals who have contributed to the completion of this thesis. First of all, I would like to acknowledge my supervisor who has inspired me in this research area, Associate Professor Fadzilah Siraj and for her remarkable guidance, integrity, generous support and comments as well as her genuineness knowledge and experience throughout the process of writing this thesis.

I would also like to thank my evaluator, Associate Professor Azizi Zakaria, for the impressive comments, suggestions and ideas in evaluating my presentation and thesis.

To a very special person in my life, my beloved husband, Mr. Shamsul Amri Abdul Manap thank you very much for the boundless encouragement, strength and love that came along side towards finishing this thesis.

I am deeply grateful to my family and family in law for their support, pray and understanding. And last but not least, I want to acknowledge my friends and co-workers for their trust and concerned.

TABLE OF CONTENTS

Permission to use	i
Abstract (Bahasa Malaysia)	ii
Abstract (English)	iii
Acknowledgement	iv
List of Tables.....	vii
List of Figures	x
Appendixes	xi
Abbreviations	xii

CHAPTER 1: INTRODUCTION

1.1 Research Background	1
1.2 Overview of College of Arts and Sciences (CAS).....	5
1.3 Problem Statement	7
1.4 Research's Objectives	8
1.5 Scope	8
1.6 Research Questions	9
1.7 Significance of Study.....	9

CHAPTER 2: LITERATURE REVIEW

2.1 Data Mining in Education	11
2.2 Academic Achievement.....	15
2.3 Logistic Regression.....	26
2.4 Artificial Neural Network (ANN).....	30

CHAPTER 3: METHODOLOGY

3.1 Business Understanding	35
3.2 Data Understanding.....	35
3.3 Data Preparation	37
3.3.1 Data Description	38
3.3.2 Data Ready to be Used	39
3.4 Modeling.....	39
3.4.1 Multinomial Logistic Regression	40
3.4.2 Neural Network	41
3.5 Evaluation	46
3.6 Deployment	46

CHAPTER 4: DATA ANALYSIS AND FINDINGS

4.1	Descriptive Statistics for UUM Graduates	47
4.2	Descriptive Statistics for CAS Students	50
4.2.1	Demographic Characteristics	50
4.2.2	Academic Information	53
4.2.3	Employment After Graduation	58
4.3	Multinomial Logistic Regression	62
4.3.1	Modelling for CAS	63
4.3.2	Modeling for FTM	69
4.3.3	Modeling for FE	72
4.3.4	Modeling for FPPH	75
4.3.5	Modeling for FSKP	78
4.3.6	Modeling for FKBM	82
4.3.7	Modeling for FPSM	85
4.3.8	Modeling for FSK	88
4.4	Neural Network	92
4.4.1	Modeling for CAS (6 targets)	92
4.4.2	Modeling for FTM	99
4.4.3	Modeling for FE	104
4.4.4	Modeling for FPPH	109
4.4.5	Modeling for FSKP	113
4.4.6	Modeling for FKBM	118
4.4.7	Modeling for FPSM	123
4.4.8	Modeling for FSK	128
4.4.9	Modeling for CAS (3 targets)	132

CHAPTER 5: CONCLUSION

5.1	Discussion	139
5.2	Conclusion	140

REFERENCES	142
APPENDIX A	148
APPENDIX B	151
APPENDIX C	159

LIST OF TABLES

Table 4.1	Case processing summary for CAS	63
Table 4.2	Distribution of target	63
Table 4.3	Classification table for CAS.....	64
Table 4.4	Model Fitting Information (CAS).....	64
Table 4.5	Pseudo R-Square (CAS).....	65
Table 4.6	Likelihood Ratio Tests (CAS).....	66
Table 4.7	List of Variables Included in students' achievement dataset (CAS).....	68
Table 4.8	Classification Table (FTM).....	69
Table 4.9	Model Fitting Information (FTM).....	69
Table 4.10	Pseudo R-Square (FTM).....	70
Table 4.11	Likelihood Ratio Tests (FTM).....	70
Table 4.12	List of Variables Included in students' achievement dataset (FTM).....	71
Table 4.13	Classification Table (FE).....	72
Table 4.14	Model Fitting Information (FE).....	72
Table 4.15	Pseudo R-Square (FE).....	73
Table 4.16	Likelihood Ratio Tests (FE).....	73
Table 4.17	List of Variables Included in students' achievement dataset (FE) ..	74
Table 4.18	Classification Table (FPPH).....	75
Table 4.19	Model Fitting Information (FPPH).....	75
Table 4.20	Pseudo R-Square (FPPH).....	76
Table 4.21	Likelihood Ratio Tests (FPPH).....	76
Table 4.22	List of Variables Included in students' achievement dataset (FPPH).....	77
Table 4.23	Classification Table (FSKP).....	78
Table 4.24	Model Fitting Information (FSKP).....	78
Table 4.25	Pseudo R-Square (FSKP).....	79
Table 4.26	Likelihood Ratio Tests (FSKP).....	79
Table 4.27	List of Variables Included in students' achievement dataset (FSKP).....	81
Table 4.28	Classification Table (FKBM).....	82
Table 4.29	Model Fitting Information (FKBM).....	82
Table 4.30	Pseudo R-Square (FKBM).....	83
Table 4.31	Likelihood Ratio Tests (FKBM).....	83
Table 4.32	List of Variables Included in students' achievement dataset (FKBM).....	84
Table 4.33	Classification Table (FPSM).....	85
Table 4.34	Model Fitting Information (FPSM)	85
Table 4.35	Pseudo R-Square (FPSM).....	86
Table 4.36	Likelihood Ratio Tests (FPSM).....	86
Table 4.37	List of Variables Included in students' achievement dataset (FPSM).....	87

Table 4.38	Classification Table (FSK).....	88
Table 4.39	Model Fitting Information (FSK).....	88
Table 4.40	Pseudo R-Square (FSK).....	89
Table 4.41	Likelihood Ratio Tests (FSK).....	89
Table 4.42	List of Variables Included in students' achievement dataset (FSK).....	90
Table 4.43a	Results to determine the best epoch.....	92
Table 4.43b	Results to determine the best data allocation.....	93
Table 4.43c	Results to determine the best average hidden unit.....	94
Table 4.43d	Results to determine the best learning rate.....	95
Table 4.43e	Results to determine the best momentum rate.....	96
Table 4.44	Parameter values for neural network model with 6 targets.....	96
Table 4.45a	Results to determine the best epoch.....	99
Table 4.45b	Results to determine the best data allocation.....	100
Table 4.45c	Results to determine the best average hidden unit.....	100
Table 4.45d	Results to determine the best learning rate.....	101
Table 4.45e	Results to determine the best average learning rate.....	102
Table 4.45f	Results to determine the best learning rate.....	102
Table 4.45g	Results to determine the best average momentum rate.....	103
Table 4.46	Parameter values for FTM model.....	103
Table 4.47a	Results to determine the best epoch.....	104
Table 4.47b	Results to determine the best data allocation.....	105
Table 4.47c	Results to determine the best average hidden unit.....	106
Table 4.47d	Results to determine the best learning rate.....	107
Table 4.47e	Results to determine the best average learning rate.....	107
Table 4.47f	Results to determine the best learning rate.....	108
Table 4.48	Parameter values for FE model.....	108
Table 4.49a	Results to determine the best epoch.....	109
Table 4.49b	Results to determine the best data allocation.....	110
Table 4.49c	Results to determine the best average hidden unit.....	111
Table 4.49d	Results to determine the best learning rate.....	111
Table 4.49e	Results to determine the best average learning rate.....	112
Table 4.49f	Results to determine the best momentum rate.....	112
Table 4.50	Parameter values for FPPH model.....	114
Table 4.51a	Results to determine the best epoch.....	114
Table 4.51b	Results to determine the best data allocation.....	114
Table 4.51c	Results to determine the best average hidden unit.....	115
Table 4.51d	Results to determine the best learning rate.....	116
Table 4.51e	Results to determine the best average learning rate.....	116
Table 4.51f	Results to determine the best momentum rate.....	117
Table 4.51g	Results to determine the best average momentum rate.....	117
Table 4.52	Parameter values for FSKP model.....	118
Table 4.53a	Results to determine the best epoch.....	119
Table 4.53b	Results to determine the best data allocation.....	119
Table 4.53c	Results to determine the best average hidden unit.....	120
Table 4.53d	Results to determine the best learning rate.....	121

Table 4.53e	Results to determine the best momentum rate -----	122
Table 4.53f	Results to determine the best average momentum rate -----	122
Table 4.54	Parameter values for FKBM model -----	123
Table 4.55a	Results to determine the best epoch -----	123
Table 4.55b	Results to determine the best data allocation -----	124
Table 4.55c	Results to determine the best average hidden unit -----	125
Table 4.55d	Results to determine the best learning rate-----	126
Table 4.55e	Results to determine the best average learning rate-----	126
Table 4.55f	Results to determine the best momentum rate -----	127
Table 4.55g	Results to determine the best average momentum rate -----	127
Table 4.56	Parameter values for FPSM model -----	128
Table 4.57a	Results to determine the best epoch -----	128
Table 4.57b	Results to determine the best data allocation -----	129
Table 4.57c	Results to determine the best average hidden unit-----	130
Table 4.57d	Results to determine the best learning rate-----	131
Table 4.57e	Results to determine the best momentum rate -----	131
Table 4.58	Parameter values for FSK model -----	132
Table 4.59	New target with three (3) classes -----	132
Table 4.60a	Results to determine the best epoch -----	133
Table 4.60b	Results to determine the best data allocation -----	134
Table 4.60c	Results to determine the best hidden unit-----	135
Table 4.60d	Results to determine the best learning rate-----	135
Table 4.60e	Results to determine the best average learning rate-----	136
Table 4.60f	Results to determine the best momentum rate -----	136
Table 4.61	Parameter values for neural network model with 3 targets-----	137
Table 4.62	Comparison of accuracy between logistic regression and neural network -----	138

LIST OF FIGURES

Figure 1.1	A List of CAS's Programmes Offered by Faculty.....	5
Figure 3.1	Phases of CRISP-DM Reference Model	34
Figure 3.2	Data sample for 2006.....	36
Figure 3.3	Data sample for 2007.....	36
Figure 3.4	Data sample for 2008.....	36
Figure 3.5	The Variables Distribution for Three Different Years.....	38
Figure 3.6	Sample Data for Multinomial Logistic Regression.....	39
Figure 3.7	Sample Data for Neural Network	39
Figure 3.8	Multinomial Logistic Regression with SPSS 12.0.....	40
Figure 3.9	MLP Icon to Control the Parameter.....	42
Figure 3.10	The Architecture of Multilayer Perceptron.....	42
Figure 3.11	Data Allocation	43
Figure 3.12	Setting the Input Layer.....	44
Figure 3.13	Multilayer Perceptron illustration.....	45
Figure 3.14	Training session.....	45
Figure 4.1	The total number of UUM graduates for year 2006 to 2008.....	48
Figure 4.2	Undergraduate students for year 2006, 2007 and 2008 by faculties	48
Figure 4.3	Comparison by gender for year 2006, 2007 and 2008.....	49
Figure 4.4	Distribution between races for all UUM graduates	49
Figure 4.5	The total number of CAS students	50
Figure 4.6	Comparison of CAS students by year.....	51
Figure 4.7	Comparison by gender for CAS students	52
Figure 4.8	Distribution between gender and races	52
Figure 4.9	The percentage of programmes enrolled by students at three different years	53
Table 4.10	Distribution of entry qualification of CAS graduates	54
Figure 4.11	Distribution of students' CGPA by class	55
Figure 4.12	Distribution of CGPA and study programmes	55
Figure 4.13	CGPA distribution between gender	56
Figure 4.14	CGPA and entry qualification by gender difference.....	56
Figure 4.15	CGPA distribution among races.....	57
Figure 4.16	Distribution of students' CGPA by family income.....	58
Figure 4.17	The percentage of graduates by employment.....	59
Figure 4.18	Distribution of students employment after graduation	60
Figure 4.19	The percentage of students' CGPA and employment	60
Figure 4.20	Employed students and CGPA achievement.....	61
Figure 4.21	Distribution of employed students by study programmes	62
Figure 4.22	Multilayer Perceptron Architecture for CAS.....	97
Figure 4.23	Confusion matrix for testing result	98
Figure 4.24	Confusion matrix for training result.....	98
Figure 4.25	Confusion matrix for testing result.....	137
Figure 4.26	Confusion matrix for training result.....	138

APPENDIXES

Appendix A Data Description 148
Appendix B List of tables for data analysis 151
Appendix C Modelling and result summary 159

ABBREVIATIONS

ANN	-	Artificial Neural Network
CAS	-	College of Arts and Sciences
CGPA	-	Cumulative Grade Point Average
COB	-	College of Business
COLGIS	-	College of Law, Government and International Studies
CS	-	Computer Science
CRISP-DM	-	Cross Industry Standard Process for Data Mining
DM	-	Data Mining
FE	-	Faculty of Economics
FKBM	-	Faculty of Communication and Modern Languages
FPPH	-	Faculty of Tourism and Hospitality
FPSM	-	Faculty of Humanities and Social Development
FSK	-	Faculty of Quantitative Sciences
FSKP	-	Faculty of Cognitive Sciences and Education
FTM	-	Faculty of Information Technology
GPA	-	Grade Point Average
ICT	-	Information and Communication Technology
IPTA	-	Institut Pengajian Tinggi Awam
KDD	-	Knowledge Discovery in Databases
MUET	-	Malaysian University English Test
UUM	-	Universiti Utara Malaysia

CHAPTER 1

INTRODUCTION

This chapter consists of a study on the students' performance of College of Arts and Sciences (CAS), Universiti Utara Malaysia (UUM). Performance information is gathered from students' final semester results. Research background, problem statements, project's objectives, scope, research questions and significance of the study are highlighted in this chapter.

1.1 RESEARCH BACKGROUND

Students' performance in academic achievement is the major concerns in the universities (Fennolar, Roman, & Cuestas, 2007). The increasing of students attending university has developed the interest in identifying factors to predict academic performance. In higher education, the issues of prediction and explanation of academic performance and a study to identify the key indicators to the academic success and persistence of students are extremely important (Komarraju, Karau & Ramayah, 2007; Ervina & Md Nor, 2005).

The contents of
the thesis is for
internal user
only

REFERENCES

- Ainin, S., & Suhana, M. (2006). Student success factors: identifying key predictors. *Journal of Education for Business*, 81(6), p328-333.
- Amral, N., Ozveren, C.S. & King, D. (2007). Short term load forecasting using Multiple Linear Regression. In *42nd International Universities Power Engineering Conference (UPEC 2007)*, (pp. 1192 – 1198).
- Akey, T. M. (2006). School context, student attitudes and behaviour, and academic achievement: an exploratory analysis. New York: MDRC Publications.
- Apte, C., Liu, B., Pednault, E. P. D. & Smyth, P. (2002). Business Application of data mining. *Communication of the ACM*, 45(8), 49-53.
- Baird, L. L. (1982). The role of academic ability in high-level accomplishment and general success. *College Board Report No. 82-6*. New York: Research and Development.
- Baxter, R. A., Gawler, M., & Ang, R. (2007). Predictive model of insolvency risk for Australian corporations. In *Proceedings of the sixth Australasian conference on Data mining and analytics*, (pp.21-27). Australia.
- Belvis, A.G., Avolio, M., Spagnolo, A., Damiani, G., Sicuro, L., Cicchetti, A., et al. (2006). Factors associated with health-related quality of life: the role of social relationships among the elderly in an Italian region. *Journal of the Royal Institute of Public Health*, 122, 784-793.
- Boyle, R., Carter, J., & Clark, M. (2002). What makes them succeed? Entry, progression and graduation in Computer Science. *Journal of Further and Higher Education*, 26(1), 3-18.
- Breen, R., & Lindsay, R. (2002). Different disciplines require different motivations for student success. *Research in Higher Education*, 43(6), 693-725.
- Cardenas, C. (2000). Development and implementation of a new model for academic evaluation. In *proceedings of 30th ASEE/ IEEE Frontiers in Education Conference*, (pp. F3A/15 - F3A/20). Mexico.
- Carroll, B. R. (2006). The effects of school violence and crime on academic achievement. Retrieved January 6, 2009, from: <http://www.davidson.edu/academic/economics/Student%20Research%20Papers/Brandon%20Carroll%20paper.pdf>

- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. (2000). CRISP-DM 1.0 step –by-step data mining guide. Retrieved January 6, 2009, from: <http://www.crisp-dm.org/CRISPWP-0800.pdf>.
- Cripps, A. (1996). Using artificial neural nets to predict academic performance. In *Proceedings of ACM Symposium on Applied Computing*, (pp. 33-37). Philadelphia, P.A.
- Dalton, W. K. (1995). Assessing student learning: it can be more than a survey. *Frontiers in Education Conference Proceedings*, 1, (pp. sc1.13-2c1.16).
- Dayioglu, M. & Turut-Asik, S. (2007). Gender differences in academic performance in a large public university in Turkey. *Journal in Higher Education*, 53, 255-277.
- Delavari, N., Beikzadeh, M. R., & Phon-Amnuaisuk, S. (2005). Application of enhanced analysis model for data mining processes in higher educational system. In *6th Annual International Conference (ITHET)*, (1-6). Juan Dolio, Dominican Republic.
- Diseth, A., Pallesen, S., Hovland, A., & Larsen, S. (2006). Course experience, approaches to learning and academic achievement. *Journal in Education and Training*, 48(2), 156-167.
- Ditcher, A. & Tetley, J. (1999). Factors influencing university students' academic success: what do students and academics think?. *HERDSA Annual International Conference*, Melbourne.
- Downing, K., Sui-Wah, C., Woo-Kyung, D., Kwong, T., & Tsz-Fung, L. (2008). Measuring gender differences in cognitive functioning. *Multicultural Education & Technology Journal*, 2(1), 4-18. Retrieved March 1, 2009, from ABI/INFORM Global database. (Document ID: 1514147951).
- Economic Planning Unit (UNDP). (2005). *Malaysia achieving the millennium development goals: successes and challenges*. Malaysia: United Nations Country Team.
- Einar, M. S., & Sidsel, S. (2006). Self-concept and self-efficacy in mathematics: relation with mathematics motivation and achievement . In *Proceedings of the 7th international conference on Learning sciences (ICLS'06)*, (pp. 709-715).
- Ervina, A., & Md Nor, O. (2005). Undergraduate students' performance: the case of University of Malaya. *Quality Assurance in Education*, 13(4), 329-343.
- Fennolar, P., Roman, S., & Cuestas, P. J. (2007). University students' academic performance: An integrative conceptual framework and empirical analysis. *British Journal of Education Psychology*, 77, 873-891.

- Geiser, S. & Santelices, M. V. (2007). Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. Standardized tests as indicators of four-year college outcomes. Retrieved January 6, 2008, from: http://eschc.berkeley.edu/publications/docs/ROPS,GEISER_SANTELICES_6.12.07.pdf
- Golding, P. & McNamarah, S. (2005). Predicting academic performance in the school of computing & information technology (SCIT). *35th ASEE/IEEE Frontiers in Education Conference*. Indianapolis.
- Hijazi, S. T. & Raza Naqvi, S. M. M. (2006). Factors affecting students' performance: a case of private colleges. *Bangladesh e-Journal of Sosiologi*, 3(1), 1-10.
- Hilton, T. L. (1982). Persistence in higher education: an empirical study. *College Board Report No. 82-5*. New York: Research and Development.
- Ho, S. K., & Nurhayati, S. A. (2001). Measuring development progress: the development indicators for Malaysia. *Paper presented at the UN Workshop on Development Indicators for the ASEAN Countries*. Manila, Philippines.
- Hostetler, T. R. (1983). Predicting student success in an introductory programming course. *SIGCSE Bulletin*, 15(3), 40-49.
- Hueglin, C., & Vannotti, F. (2001). Data mining techniques to improve forecast accuracy in airline business. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, (pp. 438 - 442). California.
- Kiang, M. Y., Fisher, D. M., Fisher, S. A., & Chi, R. T. (2005). Understand corporate rationales for engaging in reverse stock splits: a data mining application. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05)*, (pp. 75a-75l)
- Kingston, E. (2008). Emotional competence and drop-out rates in higher education. *Education & Training*, 50(2), 128-139. Retrieved March 1, 2009, from ABI/INFORM Global database. (Document ID: 1464018791).
- Kinnunen, P., McCartney, R., Murphy, L., & Thomas, L. (2007). Through the eyes of instructors: a phenomenographic investigation of student success. In *Proceedings of the third international workshop on Computing education research (ICER'07)*, (pp.61-72).
- Komarraju, M., Karau, S. J., & Ramayah, T. (2007). Cross-cultural differences in the academic motivation of university students in Malaysia and the United States. *North American Journal of Psychology*, 9(2), 275-293.

- Li, K. (2004). Application of various modeling techniques to analyze a housing condition survey dataset. *IEEE International Conference on Systems, Man and Cybernetics*, (pp. 409-414).
- Luan, J. (2001). Data mining as driven by knowledge management in higher education- persistence clustering and prediction. *SPSS Public Conference, UCSF*.
- Luan, J. (2004). *Data mining application in higher education*. SPSS Inc.
- McKenna, A., Mongia, L. & Agogino, A. (1998). Capturing student's teamwork and open-ended design performance in an undergraduate multimedia engineering design class. *28th Annual Frontiers in Education Conference (FIE'98)*, (pp. 264-269).
- McKenzie, K., Gow, K., & Schweitzer, R. (2004). Exploring first-year academic achievement through structural equation modelling. *Higher Education Research & Development*, 23(1), 95-112.
- McKenzie, K. & Schweitzer, R. (2001). Who succeeds at university? Factors predicting academic performance in first year Australian university students. *Journal of Higher Education Research & Development*, 20(1), 21-33.
- Mundfrom, D. J., Perrett, J. J., Schaffer, J., Piccone, A. & Roozeboom, M. (2006). Bonferroni adjust in tests for regression coefficients. *Multiple Linear Regression Viewpoints*, 32(1), 1-6.
- Naik, B., & Ragothaman, S. (2004). Using neural network to predict MBA student success. *College Student Journal*, 38(1), 1-4.
- National Higher Education Action Plan 2007-2010. (2007). *Triggering Higher Education Transformation*. Ministry of Higher Education Malaysia.
- Nguyen, T. N., Janeczek, P. & Haddawy, P. (2007). A comparative analysis of techniques for predicting academic performance. In *37th annual Frontiers in education conference - global engineering: knowledge without borders, opportunities without passports (FIE '07)*, (pp. T2G-7 - T2G-12).
- Nissen, A. (2005). Neural network made simple. Retrieved March 6, 2009, from: [http://www. Software20.org](http://www.Software20.org) .
- Noor Azina, I., & Azmah, O. (2006). Comparing university academic performances of HSC students at the three art- based faculties. *International Education Journal*, 7(5), 668-675.
- Nordaliela, M. R., Zaidah, I., & Roziah, M. J. (2008). Predicting Students' Academic Achievement: Comparison between Logistic Regression, Artificial Neural Network, and Neuro-Fuzzy. *International Symposium on Information Technology (ITSim 2008)*, 1, (pp.1 – 6).

- Ogor, E. N. (2007). Student academic performance monitoring and evaluation using data mining techniques. *Electronics, Robotics and Automotive Mechanics Conference (CERMA 2007)*, (pp. 354-359).
- Oladokun, V. O., Adebajo, A. T., & Charles-Owaba, O. E. (2008). Predicting students' academic performance using artificial neural network: a case study of an engineering course. *The Pacific Journal of Science and Technology*, 6(1), 72-79.
- Parent, D. W., & Rio-Parent, L. D. (2008). Introducing TCAD in a graduate level device physics course. *IEEE Transaction on Education*, 51(3), 331-335.
- Pedrosa, R. H. L., Dachs, J. N. W., Maia, R. P., Andrade, C. Y., & Carvalho, B. S.. (2007). Academic performance, students' background and affirmative action at a Brazilian University. *Higher Education Management & Policy*, 20 (1), p67-86.
- Pimentel, E.P. & Nizam, O. (2005). Towards a model for organizing and measuring knowledge upgrade in education with data mining. In *IEEE International Conference on Information Reuse and Integration (IRI -2005)*, (pp. 56-60). Nevada, USA.
- Pomykalski, J. J., Dion, P., Brock, J. L. (2008). A Structural equation model for predicting business student performance. *Journal of Education for Business*, 83(3), 159-163. Retrieved March 1, 2009, from ABI/INFORM Global database. (Document ID: 1517979621).
- Price, B. A., & Randall, C. H. (2004). Technology-based degrees: do faculty and students agree on necessary skills for success in these programs?. In *Proceedings of the 5th conference on Information technology education (CITC5 '04)*, (pp. 235-240). UT, USA.
- Quinn, A. S. (1999). Neural connection 2.1. Retrieved April, 6 2009, from: www.und.edu/instruct/aquinn/neural_connection.pdf
- Salazar, A., Gosalbez, J., Bosch, I., Miralles, R. & Vergara, L. (2004). A case study of knowledge discovery on academic achievement student desertation and student retention. In *proceedings of International Conference on Information Technology: Research and Education (ITRE 2004)*, (pp. 150-154).
- Scott, C. L., & Homant, R. J. (2007). The professional mentor program plus: an academic success and retention tool for adult learners. *Journal of College Student Retention*, 9(1), 61-73.
- Smith, F. (2004). "It's not all about grades": Accounting for gendered degree results in geography at Brunel university. *Journal of Geography in Higher Education*, 28(2), 167-178.

- Sulaiman, K. M. (1987). The relationship of selected predictive variables to the academic success of students from the United Arab Emirates at American colleges and universities. (Doctoral dissertation, University of George Washington, 1987). *Dissertation Abstracts International*, 47 (12-A, Pt 1), 4309.
- Swope, K. J., & Schmitt, P. M. (2006). The performance of economics graduates over the entire curriculum: the determinants of success. *Journal of Economic Education*, 37 (4), 387-394.
- Taafe, M., & Cunningham, E. G. (2005). Retrieved March 6, 2009, from: www.aare.edu.au/05pap/taa05428.pdf
- Tan, P., Steinbach, M. & Kumar, V. (2006). *Introduction to data mining*. Boston: Addison Wesley.
- Taylor, H. G. & Mounfield, L. C. (1989). The effect of high school computer science, gender and work on success in college computer science. In *Proceedings of the twentieth SIGCSE technical symposium on Computer science education (SIGCSE '89)*, (pp. 195-198).
- Tissera, W. M. R., Athauda, R. I., & Fernando, H. C. (2006). Discovery of strongly related subjects in the undergraduate syllabi using data mining. *International Conference on Information and Automation (ICIA '06)*, (pp. 57-62).
- Toh, K. S. (2009). Faktor mempengaruhi akademik lelaki dan wanita. Retrieved March 6, 2009, from: <http://www.scribd.com/doc/12521518/Faktorfaktor-mempengaruhi-pencapaian-akademik-lelaki-dan-perempuan>
- Vaez, M., & Laflamme, L. (2008). Experienced stress, psychological symptoms, self-rated health and academic achievement: a longitudinal study of Swedish university students. *Social Behavior and Personality*, 36 (2), 183-196
- Vranic, M., Pintar, D. & Skocir, Z. (2007). The use of data mining in education environment. In *9th International Conference on Telecommunications (ConTel 2007)*, (pp. 243-250). Croatia.
- Zaidah, I. & Daliela, R. (2007). Predicting student's academic performance: comparing artificial neural network, decision tree and linear regression. *21st Annual SAS Malaysia Forum*, (pp.1-6). Kuala Lumpur.